

### Bioinformatics for Cancer Epidemiologic Research

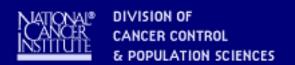
Deborah Winn

Division of Cancer Control and
Population Sciences Health
Informatics Initiative





## Primary goal of the NCI DCCPS health informatics initiative

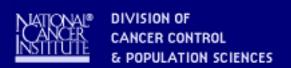


- To transform research in cancer control and population sciences by using the benefits of health informatics to
  - Improve the efficiency and reduce the difficulties of conducting research
  - Increase cross-disciplinary and collaborative research, especially through the use of standards and common platforms
  - Create opportunities for data and other linkages
  - Increase capacity for research and the use of research to understand mechanisms of cancer causation, monitor the burden of cancer, develop interventions to prevent cancer, detect it earlier, and reduce its consequences, and inform policy decisions





## Trends influencing the need for a bioinformatics initiative

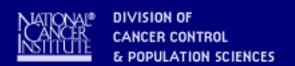


- NIH data sharing policy
- © Collaboration occurring among scientists who may "speak different languages"
- © Growth of consortia
- Need for "cross-walks" across existing studies
- © Epidemiology studies involve data from many content domain types: genotypes, biochemistries, clinical record, questionnaire data





# Trends influencing the need for a bioinformatics initiative (cont.)

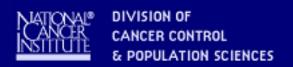


- © Constant need to do research faster, better, cheaper
- Need to translate from biology to epidemiology to prevention and back
- NIH roadmap: Bioinformatics and Computational Biology
- Strong NCI bioinformatics development efforts in other areas, e.g., HAPMAP





# Scope of DCCPS Health Informatics Initiative



- Surveillance systems
- Behavioral research
- Intervention research
- © Cancer outcomes and quality of care
- © Cancer survivorship







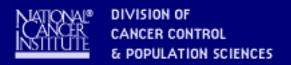
### **NCI** partners

- Olivision of Cancer Control and Population Sciences (DCCPS)(Extramural)
- Olivision of Cancer Epidemiology and Genetics (DCEG)(Intramural)
- © Center for Bioinformatics (NCICB)





### DCCPS and DCEG joint activities

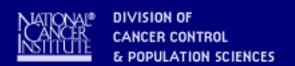


- © Development of a Vocabulary Enterprise System for population sciences and cancer control concepts
- A population sciences and cancer control "workspace" within the NCICB caBIG project
- Models and applications, e.g., risk prediction workshop





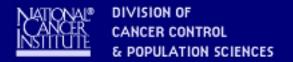
## Selected major initiatives of the NCI Center for Bioinformatics



- @caBIG
- © Enterprise vocabulary systems







### caBIG

- O An initiative of the NCICB has as its purpose a common, extensible informatics platform that integrates diverse data types and supports interoperable analytic tools. Will allow research groups to tap into the rich collection of emerging cancer research data while supporting their individual investigations. Workspaces within the caBIG initiative are areas or virtual environments where caBIG pilot project activities will be grouped.
- © DCCPS and DCEG are working towards implementation of a population sciences and cancer control workspace within caBIG to enhance both the research underway or supported by the two Divisions as well as that of other scientists engaged in cancer research





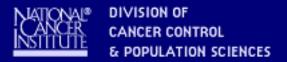
### Enterprise vocabulary system

The NCI EVS is set of services and resources that address NCI's needs for controlled vocabulary. It includes activities related to common data elements and the NCI Thesaurus and Metathesaurus, which is a biomedical thesaurus created specifically to meet the needs of the NCI, is produced by the NCI EVS project.





### Common Data Elements

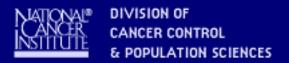


The Common Data Elements (CDE)
Dictionary contains a repository of detailed information pertaining to each standard element used to collect essential data related to cancer research





## Priority content domains

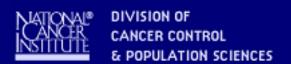


- Tobacco use
- Alcohol use
- Operation
  Diet
- Physical activity
- Body mass
- Medications
- Reproductive factors
- © Cancer screening





# Nearly completed Common Data Elements

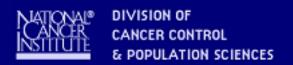


- @Age
- ©Sex
- Race





## Value of common data elements and an established caBIG workspace

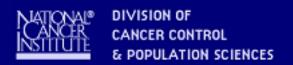


- Have available a repository of information about the data in studies
- Can facilitate creation of questionnaires and forms
- Can compare data elements from different studies
- Can cross-walk between studies if they have been added to the Enterprise Vocabulary System





## Value of common data elements & an established caBIG workspace (cont.)

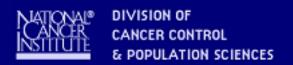


- Can benefit from common data elements and thesauruses from other disciplines
- Helps ensure that other disciplines can benefit from our knowledge about appropriate ways to collect information related to population sciences and cancer control





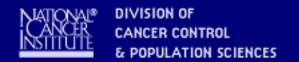
# Cancer Family Registries Informatics Center



- Contract to be let soon
- Intended to make all Cancer Family Registry informatics practices consistent with NCI Center for Bioinformatics standards
- © Converts all of existing cancer family registry questionnaires and forms (e.g., family history information) into common data elements
- Will serve as a model for related efforts







### Uncle Sam wants you!



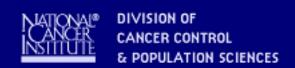
Uncle Sam wants your questionnaires!

- In an advisory capacity
- On common data elements working groups

- Electronic format if possible
- Any cross-walking you have done



# Some useful NCI Center for Bioinformatics websites



- Ohome page <a href="http://ncicb.nci.nih.gov/">http://ncicb.nci.nih.gov/</a>
- © Cancer data standards repository http://ncicb.nci.nih.gov/core/caDSR
- © Common data elements browser:
  <a href="http://ncicb.nci.nih.gov/CDEBrowse">http://ncicb.nci.nih.gov/CDEBrowse</a>
  <a href="mailto:r/">r/</a> (I used smok\* as my search word)
- ONCI metathesaurus page: <a href="http://ncimeta.nci.nih.gov/indexMeta">http://ncimeta.nci.nih.gov/indexMeta</a> <a href="phrase.html">phrase.html</a> Click on "browse" to start.





# Epidemiology and Genetics Research Program contact



Deborah M. Winn, Ph.D., Chief Clinical & Genetic Epidemiology Research **Branch Epidemiology and Genetics Research Program Division of Cancer Control and Population** Sciences Executive Plaza North, Room 5134, MSC 7393 6130 Executive Boulevard Bethesda, MD 20892-7393 (Regular mail) Rockville, MD 20852 301-594-9499 (Voice) 301-496-9600 (Main group number) 301-435-5477 (Fax)

















**DIVISION OF** CANCER CONTROL & POPULATION SCIENCES

	DE Browse	er			CDE Home Fo	FormBuilder	r Help
					Data I	Element S	Search
Note: Enter/select search crit information on CDEBrowser.	iteria and click search button to initia	ate search, <sup>1</sup>	The wildcard (	character is *. Click	the Help button abo	ove for more	
Search Fo			Al	ternate Name:			
Search Field(s)	ALL Broforred Name			Name Type(s):	,	in E	
Permissible Value	a:						
Value Domain	1:			Public ID:			
Data Element Concept	t /		3	Classification:			€ Cle
Version	n: Latest Version · All Vers	sions O		Context Use:	Owned By/Use	d By 🔻	
Workflow Status	ALL APPRVD FOR TRIAL USE CMTE APPROVED CMTE SUBMTD		Regis	stration Status:	ALL Application Candidate Qualified		
Download Data Eleme	Search Data Elem		Clear ments as XI	New Se	arch		
Add to CDE Cart					8	1 - 21 of 2	21 🔻 👂
□ Long Name	Document Text	Owned By	Used By Context	The second secon	Workflow Status	D. blis	Version
Smoked at least 100 cigarette (5 packs) during lifetime	Smoked at least 100 cigarette (5 packs) during lifetime	CTEP	SPOREs	Qualified	APPRVD FOR TRIAL USE	2001713	3.0
Smoking Current Use Daily Frequency	Number of Cigarettes per day	CTEP		Qualified	RELEASED	2179794	1.0
Smoking Current Use Ind-2	Does the patient currently smoke	CTEP		Qualified	RELEASED	2180053	1.0











Data Element Concept

Permissible Values

Classifications

Usage

Data Element Derivation

### **Data Element Details**

Public ID: 2001713

Preferred Name: SMOKED AT LEAST 100 CIGARETTE

Long Name: Smoked at least 100 cigarette (5 packs) during lifetime

Document Text: Smoked at least 100 cigarette (5 packs) during lifetime

Definition: Smoked at least 100 cigarette (5 packs) during lifetime

Value Domain: SMOKED\_AT\_LEAST\_100\_CIGARETTE\_

Data Element Concept: SMOKED\_AT\_LEAST\_100\_CIGARETTE\_

Context: CTEP

Workflow Status: APPRVD FOR TRIAL USE

Version: 3.0

Origin:

Registration Status: Qualified

### Reference Documents

Document Name	Document Type	Document Text	URL
CRF Text	LONG_NAME	Smoked at least 100 cigarette (5 packs) during lifetime	
CRF Text	HISTORIC SHORT CDE NAME	Smoked at least 100 cigarette (5 packs) during lifetime	

### Designations

Name	Туре	Context	Language
SMOKED_AT_LEAST_100_CIGARETTE_	USED_BY	SPOREs	







### **DIVISION OF** CANCER CONTROL & POPULATION SCIENCES







Data Element V Data Element Concept Permissible Values Classifications V Usage V Data Element Derivation

### Selected Data Element

Public ID:	2001713
Preferred Name:	SMOKED_AT_LEAST_100_CIGARETTE_
Long Name:	Smoked at least 100 cigarette (5 packs) during lifetime
Document Text:	Smoked at least 100 cigarette (5 packs) during lifetime
Definition:	Smoked at least 100 cigarette (5 packs) during lifetime
Workflow Status:	APPRVD FOR TRIAL USE
Version:	3.0

### Value Domain Details

D 111 ID	2010000
Public ID:	
Preferred Name:	SMOKED_AT_LEAST_100_CIGARETTE_
Long Name:	Smoked at least 100 cigarette (5 packs) during lifetime
Definition:	Smoked at least 100 cigarette (5 packs) during lifetime
Workflow Status:	DRAFT NEW
Version:	3.0
Datatype:	ALPHANUMERIC
Unit of Measure:	
Display Format:	
Maximum Length:	255
Minimum Length:	
Decimal Place:	
High Value:	
Low Value:	
Value Domain Type:	
Conceptual Domain Preferred Name:	СТЕР
Conceptual Domain Context Name:	
Conceptual Domain Version:	2.31
Origin:	

### Permissible Values

Value	Value Meaning	Value Meaning Description
No	NO	NO
Yes, but quit	YES, BUT QUIT	
Yes, currently smoke	YES, CURRENTLY SMOKE	

user: Public User













Data Element Concept

Permissible Values

Classifications

<u>Usage</u>

Data Element Derivation

### Selected Data Element

Public ID: 2001713

Preferred Name: SMOKED\_AT\_LEAST\_100\_CIGARETTE\_

Long Name: Smoked at least 100 cigarette (5 packs) during lifetime

Document Text: Smoked at least 100 cigarette (5 packs) during lifetime

Definition: Smoked at least 100 cigarette (5 packs) during lifetime

Workflow Status: APPRVD FOR TRIAL USE

Version: 3.0

There are no classifications for the selected CDE.

### Classifications

\*CS:Classification Scheme | CSI:Classification Scheme Item

CS\* Preferred Name

CS\* Definition

CSI\* Name

CSI\* Type

user: Public User

Version 2.1.











Data Element Concept

Permissible Values

Classifications

Usage

Data Element Derivation

### Selected Data Element

Public ID:	2001713
Preferred Name:	SMOKED_AT_LEAST_100_CIGARETTE_
Long Name:	Smoked at least 100 cigarette (5 packs) during lifetime
Document Text:	Smoked at least 100 cigarette (5 packs) during lifetime
Definition:	Smoked at least 100 cigarette (5 packs) during lifetime
Workflow Status:	APPRVD FOR TRIAL USE
Version:	3.0

### Usage

Protocol Number	Lead Org	Used By	Usage Type
lloprost	LCBCC	SMOKING HISTORY	CRF
RTOG P-0011	RTOG	A5 Radiation Therapy Oncology Group Demographic Data Form - Version 2	CRF
RTOG-0320	RTOG	A5 RTOG Phase III - Multiple Brain Mets Demographic Data Form - version 2	CRF
RTOG0214	RTOG	Radiation Therapy Oncology Group Demographic Data Form - Version 2	CRF













Data Element Concept

Permissible Values

Classifications

<u>Usage</u>

**Data Element Derivation** 

### Selected Data Element

Public ID: 2001713

Preferred Name: SMOKED\_AT\_LEAST\_100\_CIGARETTE\_

Long Name: Smoked at least 100 cigarette (5 packs) during lifetime

Document Text: Smoked at least 100 cigarette (5 packs) during lifetime

Definition: Smoked at least 100 cigarette (5 packs) during lifetime

Workflow Status: APPRVD FOR TRIAL USE

Version: 3.0

### **Data Element Derivation Details**

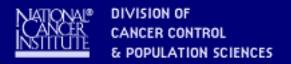
Selected CDE is not a derived data element.

user: Public User

Version 2.1.1















epidemiological

About

Browse



The basic search is enabled. Click on the "Advanced" tab to customize your search criteria. You can mouse-over each advanced search item for help in utilizing the advanced features.

Research Activity

Idea or Concept

Idea or Concept

Qualitative Concept

Occupational Activity

Research Activity

Research Activity

Research Activity; Intellectual Product

Research Activity; Intellectual Product

### **Matching Concepts**

Epidemiological Approaches

Epidemiological Factors

Epidemiological Methods Copyright

**New Term** 

Sources

User's Guide

epidemiological bias (Bias (Epidemiology))

epidemiological indicators

Epidemiological Monitoring

Epidemiological Studies (Epidemiologic Studies) Epidemiological Studies (Epidemiology Research)

Epidemiological Techniques (Epidemiological Methods)

If you did not find the concept you're looking for, and you're sure your spelling is correct, please click:

Suggest New Term











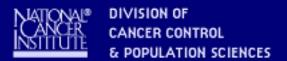
### DIVISION OF CANCER CONTROL & POPULATION SCIENCES

NATIONAL® INSTITUTE		EVS ENTERPRISE VOCABULARY SERVER	Basic	Advanced
epidemiological	Search	The basic search is enabled, Click on the "A You can mouse-over each advanced search	dvanced" tab to customize your search criteria. item for help in utilizing the advanced features.	
	Concept   De	efinitions   Synonyms   Sources   Broader Concepts   I <u>Procedures   Laboratory   Diagnosis   Open NCI</u>		ications
About	3	CL003599: Epidemiological F	actors	2
Browse	Idea or Conce	pt		
Copyright		Epidemiological Factors Defin	nitions	2
New Term	Source	Def	finition	
Copyright	NCI	those things which contribute to the establishment and tra	ansmission of disease.	
New Term				
Copyright		Epidemiological Factors Syno	onyms	?
New Term	Epidemiologica	al Factors		
Copyright		Epidemiological Factors Sou	urces	2
New Term	NCI			
Copyright	NCI			
New Term	View neighborho	od in NCI OK	Table View Of Sources	
Copyright		Concepts More General Than Epidemi	iological Factors	2
New Term	(inverse_isa)	Conceptual Entities		
Copyright		Concepts More Specific Than Epidemi	ological Factors	7
New Term			, , , , , , , , , , , , , , , , , , ,	
Copyright	(isa) (isa)	Chemical Agents Dietary Factors		
New Term	(isa)	Endogenous Factors		
0	(isa)	Epidemiology, Age, Sex		
Copyright	(isa)	Epidemiology, Geographic Location Epidemiology, Occupation		
New Term	(isa) (isa)	Epidemiology, Occupation  Epidemiology, Race/Ethnic Origin		
Copyright	(isa)	Epidemiology, Tobacco, Alcohol, Drugs		
New Term	(isa)	Exogenous Factors		
Sources	(isa) (isa)	Knowledge, Attitudes, Behaviors Lifestyle Factors, Unspecified		
User's Guide	(isa)	Societal Factors		
		Concepts Related To Epidemiologi	ical Factors	2
		• •		
Center		Medications		7
for		Procedures		7
Bioinformatics		rrocedures		Œ.
FIRSTGOV Year Free Click in the U.S. Government		Laboratory		2
		Diagnosis		2







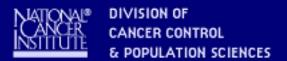


NATIONAL® CÂNCER INSTITUTE	_	EVS ENTERPRISE VOCABULARY SERVER Basic	Advanced
epidemiological	Search	The basic search is enabled. Click on the "Advanced" tab to customize your search criteria. You can mouse-over each advanced search item for help in utilizing the advanced features.	
	Concept   I	Definitions   Synonyms   Sources   Broader Concepts   Narrower Concepts   Related Concepts   Merocedures   Laboratory   Diagnosis   Open NCI Hierarchy   View Hierarchy Location	dications
About		CL000367: Epidemiology, Tobacco, Alcohol, Drugs	2
Browse	Research Ac	tivity	
Copyright		Epidemiology, Tobacco, Alcohol, Drugs Definitions	2
New Term	Source	Definition	
Copyright	NCI	consideration of tobacco, alcohol, or drug use as variables in disease transmission and control.	
New Term		ovalidation of the organization of the organiz	
Copyright		Epidemiology, Tobacco, Alcohol, Drugs Synonyms	2
New Term	Epidemiology	r, Tobacco, Alcohol, Drugs	
Conveight		T. 'l., 'l., T.L., Al. I.I.D., C.	7
Copyright New Term		Epidemiology, Tobacco, Alcohol, Drugs Sources	LCI
New Term	NCI		
Copyright New Term	View neighborh	ood in NCI OK Table View Of Sources	
New Term		SCHOOL SC	
Copyright		Concepts More General Than Epidemiology, Tobacco, Alcohol, Drugs	2
New Term	(inverse_isa)	Epidemiological Factors	
Sources		Concepts More Specific Than Epidemiology, Tobacco, Alcohol, Drugs	7
User's Guide			
		Concepts Related To Epidemiology, Tobacco, Alcohol, Drugs	2
Center		Medications	2
for		Procedures	[7]
Bioinformatics		Tioceames	-
FIRSTGOV		Laboratory	2
		Diagnosis	7





### **NCI** Metathesaurus









epidemiological



The basic search is enabled. Click on the "Advanced" tab to customize your search criteria. You can mouse-over each advanced search item for help in utilizing the advanced features.

### Sources of Epidemiology, Tobacco, Alcohol, Drugs

About

Browse

Copyright

**New Term** 

Sources

User's Guide

Source	Туре	Code	Term
NCI	PT	<u>C18774</u>	Epidemiology, Tobacco, Alcohol, Drugs
RAEB-2	PT	N023242	Epidemiology, Tobacco, Alcohol, Drugs

An explanation of the UMLS term types can be found at the NLM's documentation page here.







